

Information content of the near-field I: two-dimensional samples

Richard A. Frazin

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, Urbana, IL 61821

David G. Fischer

NASA Glenn Research Center, MS110-3, 21000 Brookpark Rd.
Cleveland, OH 44135

P. Scott Carney

Department of Electrical and Computer Engineering, The Beckman
Institute for Advanced Science and Technology, and The Coordinated
Science Laboratory, University of Illinois at Urbana-Champaign, Urbana,
IL 61821

Limits on the effective resolution of many optical near-field experiments are investigated. The results are applicable to variants of total internal reflection microscopy (TIRM), photon scanning optical microscopy (PSTM) and near-field optical scanning microscopy (NSOM) in which the sample is weakly scattering and the direction of illumination may be controlled. Analytical expressions for the variance of the estimate of the complex susceptibility of an unknown two-dimensional object as a function of spatial frequency are obtained for Gaussian and Poisson noise models and a model-independent measure is examined. The results are used to explore the transition from near-zone to far-zone detection. It is demonstrated that the information content of the measurements made at a distance of even one wavelength away from the sample is already not much different from the information content of the far-field.

© 2004 Optical Society of America

OCIS codes: 110.0180, 180.5810, 290.3200, 070.6020, 110.3000.

1. Introduction

Near-field microscopy^{1,2} has emerged as an increasingly important imaging technology in the last two decades. The principal advantage of near-field optical methods is the ability to provide images with resolution on scales much smaller than the Rayleigh-Abbe resolution limit.³ Synge⁴ first suggested the basic method in which a thin sample is illuminated through a subwavelength aperture that is scanned very close (less than a wavelength away) to the sample while the amount of transmitted

light is recorded as a function of position. This method is now known as near-field scanning optical microscopy (NSOM),⁵⁻⁸ and is often practiced by illuminating the sample and collecting the light with a subwavelength aperture at the end of a tapered optical fiber. Photon scanning tunneling microscopy (PSTM) is a related method in which the sample is illuminated by an evanescent wave created by total internal reflection at a prism face and the subsequently scattered light is collected by a tapered fiber as in NSOM.

Total internal reflection microscopy (TIRM) also takes advantage of sample illumination by evanescent waves at a prism face. Unlike NSOM or PSTM, in TIRM the scattered radiation is measured in the far-field. TIRM has been used primarily as a surface inspection technique,^{9,10} however, recently there has been interest in taking advantage of the high spatial frequency content of the evanescent wave¹¹ to perform direct imaging with transverse resolution surpassing the Rayleigh-Abbe limit.¹²

Quantitative interpretation of NSOM, PSTM and TIRM images is sometimes difficult. Under certain simplifying assumptions such as homogeneity of the material dielectric susceptibility,⁶ the measurements may be related to topography. Under more general conditions, a solution of the linearized scattering problem¹³⁻¹⁶ may be employed to relate near-field optical measurements to object structure and composition. An important question remains: What is the resolution that a given experiment can hope to achieve?

This paper addresses the question of resolution through an analysis of the variance in the Fourier components of the estimate of the sample. The linearized inverse scattering solution¹³⁻¹⁶ is applied to the case of a thin, weakly scattering sample and

an analytic expression for the error covariance matrix of the estimate of the sample susceptibility is obtained for Gaussian and Poisson noise models. Furthermore, a noise-model independent bound on the estimate error is obtained. These formulae are then applied to PSTM for increasing distances between the sample and measurement plane until the experiment is essentially TIRM. It is seen that in the near-zone one may surpass the usual diffraction limit with the bounds being set by the signal-to-noise ratio. It is shown that the information content in the plane that is only a wavelength away from the sample is already quite similar to that of the far-field, i.e., one obtains the usual Ewald sphere of reflection.³

2. Scattering model

Consider an experiment in which a monochromatic field is incident on a dielectric medium with susceptibility $\eta(\mathbf{r})$. The field incident on the sample is taken to be a plane wave. One half space, taken to be $z \geq 0$, has the vacuum index of refraction while the $z < 0$ half space has an index of refraction n . Only nonmagnetic materials are considered and so it is sufficient to treat only the electric field \mathbf{E} . The field satisfies the equation

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - n^2(z)k_0^2 \mathbf{E}(\mathbf{r}) = 4\pi k_0^2 \eta(\mathbf{r}) \mathbf{E}(\mathbf{r}), \quad (1)$$

where k_0 is the free space wave number and $n(z)$ is the z dependent background index of refraction as described above.

The sample is assumed to be of constant thickness Δz , where Δz is much less than a wavelength, and to depend only on the transverse spatial coordinate $\boldsymbol{\rho}$, where

$\mathbf{r} = (\boldsymbol{\rho}, z)$ so

$$\begin{aligned}\eta(\mathbf{r}) &= \eta(\boldsymbol{\rho}) \quad \text{for } 0 \leq z < \Delta z \\ &0 \quad \text{for } z \geq \Delta z .\end{aligned}$$

The field is taken to consist of two parts, $\mathbf{E} = \mathbf{E}^i + \mathbf{E}^s$, where \mathbf{E}^i is the incident field satisfying Eq. (1) with $\eta(\mathbf{r}) \equiv 0$. The scattered field \mathbf{E}^s arises as a result of $\eta(\mathbf{r}) \neq 0$. The incident field is a (possibly evanescent) plane wave

$$\mathbf{E}^i(\mathbf{r}) = \mathbf{e}^i \exp [i\mathbf{k}(\mathbf{q}_i) \cdot \mathbf{r}] , \quad (2)$$

where \mathbf{q}_i is the transverse wave vector of the incident wave and \mathbf{e}_i is the complex field vector such that $|\mathbf{e}_i|^2$ has units of energy/(time \cdot area). The wave vector \mathbf{k} is specified by the transverse components \mathbf{q} such that

$$\mathbf{k}(\mathbf{q}) = (\mathbf{q}, k_z(\mathbf{q})) , \text{ and} \quad (3)$$

$$k_z(\mathbf{q}) = \sqrt{k_0^2 - |\mathbf{q}|^2} . \quad (4)$$

The modes for which $|\mathbf{q}| < k_0$ are homogeneous, or propagating, plane waves. When $|\mathbf{q}| > k_0$ the plane wave is evanescent, decaying exponentially with increasing values of z . These waves are superoscillatory in the transverse plane and thus provide a means to probe the high spatial frequency structure of the sample. Keeping only the first term in the Born series,³ the scattered field is given by the expression:¹

$$E_\alpha^s(\mathbf{r}) = k_0^2 \int d^3\mathbf{r}' G_{\alpha\beta}(\mathbf{r}, \mathbf{r}') e_\beta^i \exp [i\mathbf{q}_i \cdot \boldsymbol{\rho}' + k_z(\mathbf{q}_i)z'] \eta(\mathbf{r}') , \quad (5)$$

where G is the 1/2-space Green's tensor,¹ and $\boldsymbol{\rho}'$ is equal to the transverse components of \mathbf{r}' .

The measurements are made in the plane $z = z_d$, which is illuminated by a reference wave with polarization \mathbf{e}^r

$$\mathbf{E}^r(\mathbf{r}) = \mathbf{e}^r \exp [i\mathbf{k}(\mathbf{q}_r) \cdot \mathbf{r}] . \quad (6)$$

Note that in the NSOM and PSTM imaging modalities, the incident field also plays the role of the reference field. The total field in the detector plane is given by the superposition of the reference and scattered fields, and the resulting intensity, called the *total intensity*, is:

$$I_T(\boldsymbol{\rho}, z_d) = E_\alpha^r \cdot E_\alpha^{r*} + E_\alpha^s \cdot E_\alpha^{s*} + E_\alpha^s \cdot E_\alpha^{r*} + E_\alpha^r \cdot E_\alpha^{s*} , \quad (7)$$

where the spatial arguments of the fields have been suppressed and the $*$ notation indicates that the complex conjugate is to be taken. The first term is constant at fixed z_d . The second term is necessarily non-linear in the susceptibility. Under weak scattering conditions, it is negligible compared to the third and fourth terms and will be disregarded hereafter. However, if the reference field is evanescent the scattering term becomes more important and eventually dominates as z_d increases. The third and fourth terms are conjugates that carry image information, which together will be referred to as the *holographic intensity*:

$$I(\boldsymbol{\rho}, z_d) \equiv E_\alpha^s \cdot E_\alpha^{r*} + E_\alpha^r \cdot E_\alpha^{s*} . \quad (8)$$

It should be noted that $I(\boldsymbol{\rho}, z_d)$ may be positive or negative.

Determination of η is facilitated by taking the 2D (transverse) Fourier transform of the holographic intensity $I(\boldsymbol{\rho}, z_d)$, i.e.,

$$\tilde{I}(\mathbf{Q}, z_d) \equiv \frac{1}{2\pi} \int d^2\rho I(\boldsymbol{\rho}, z_d) \exp[i\mathbf{Q} \cdot \boldsymbol{\rho}] . \quad (9)$$

Substituting Eqs. (5) and (6) into (8) yields

$$\tilde{I}(\mathbf{Q}, z_d) = [H(\mathbf{Q}, z_d)\tilde{\eta}(\mathbf{Q} + \mathbf{q}_i - \mathbf{q}_r) + H^*(-\mathbf{Q}, z_d)\tilde{\eta}^*(-\mathbf{Q} + \mathbf{q}_i - \mathbf{q}_r)] , \quad (10)$$

where $\tilde{\eta}^*$ represents complex conjugate of the Fourier transform of η .

The expression for $H(\mathbf{Q}, z_d)$ is:

$$H(\mathbf{Q}, z_d) = \frac{ik_0^2}{2\pi} e_{\beta}^{r*} e_{\alpha}^i \frac{\Gamma_{\alpha\beta}(\mathbf{q}_r - \mathbf{Q})}{k_z(\mathbf{q}_r - \mathbf{Q})} \exp\{i[k_z(\mathbf{q}_r - \mathbf{Q}) - k_z^*(\mathbf{q}_r)]z_d\} . \quad (11)$$

The tensor $\Gamma_{\gamma\delta}$ contains the integration over z' and is given in the appendix. It is worthwhile to note the following properties of Eq. (10):

1. When $|\mathbf{q}_r - \mathbf{Q}| > k_0$ or $|\mathbf{q}_r| > k_0$, the value of $H(\mathbf{Q}, z_d)$ decreases exponentially with z_d . Similarly, when $|\mathbf{q}_r + \mathbf{Q}| > k_0$ or $|\mathbf{q}_r| > k_0$, the value of $H^*(-\mathbf{Q}, z_d)$ decreases exponentially with z_d .
2. The arguments of both $\tilde{\eta}$ and $\tilde{\eta}^*$ are shifted by $\mathbf{q}_i - \mathbf{q}_r$. This allows the possibility for some of the super-oscillatory components of $\tilde{\eta}(\mathbf{Q})$ with $|\mathbf{Q}| > k_0$ to be carried into the far field via homogenous waves.

At this point one may generalize the result of Ref. 17, where the scattered fields in TIRM experiments are related to object structure, to include the effect of the holographic measurement. To understand the results it is helpful first to consider a heuristic argument. The objective is to determine the maximum value of $|\mathbf{Q}|$ such that the value of $\tilde{\eta}(\mathbf{Q})$ may be determined from measurements in the far-zone. Consider only the first term in Eq. (10). In order to avoid exponentially decreasing $H(\mathbf{Q}, z_d)$ with z_d , \mathbf{q}_r and \mathbf{Q} must satisfy the conditions $|\mathbf{Q} - \mathbf{q}_r| \leq k_0$ and $|\mathbf{q}_r| \leq k_0$. Recalling

that for the TIRM and PSTM imaging modalities $|\mathbf{q}_i| \leq nk_0$ (where n is the prism index of refraction), and setting $\mathbf{Q} - \mathbf{q}_r = k_0\hat{\mathbf{x}}$ where $\hat{\mathbf{x}}$ is the unit vector in the x direction, it may be seen that $|\mathbf{Q} + \mathbf{q}_i - \mathbf{q}_r|$ is maximized when $\mathbf{q}_i = nk_0\hat{\mathbf{x}}$. Thus, $\mathbf{Q} + \mathbf{q}_i - \mathbf{q}_r = (n+1)k_0\hat{\mathbf{x}}$. Since the x direction was chosen arbitrarily, $\hat{\mathbf{x}}$ may be replaced by any unit vector in the transverse plane. Therefore, in principle, $\tilde{\eta}(\mathbf{Q})$ may be determined when $|\mathbf{Q}| \leq (n+1)k_0$. This is enough information to construct a low-pass filtered version of the object. For comparison, the Fourier transform of the expression given by So¹² for the point spread function of a confocal microscope leads to a bandlimit of $|\mathbf{Q}| \leq \text{NA}k_0$, where NA is the numerical aperture (recall that $\text{NA} < n$).

3. Statistical model

Real measurements are made on a space-limited array. For simplicity, assume that the detection surface has an area of $L \times L$, lies in the plane $z = z_d$, and that it has $M \times M$ pixels that give the total intensity at points $\{\boldsymbol{\rho}_{mn}\}$. Since the H functions are proportional to $\exp(-|\mathbf{Q}|z_d)$ when $|\mathbf{Q}| \gg \mathbf{q}_r$ and $|\mathbf{Q}| \gg \mathbf{q}_i$, the intensity in the detector plane is effectively band-limited. Therefore Shannon's theorem applies and the intensity in the detector plane can be represented by a discrete set of uniformly spaced samples.¹⁸ The pixels are assumed to be small enough so that the intensity $I_T(\boldsymbol{\rho}, z_d)$ in Eq. (7) is essentially constant throughout each pixel (i.e., the detector does not introduce significant low-pass spatial filtering). It is also assumed that the detector has enough surface area so that nearly all of the scattered light is collected. Henceforth, M is taken to be an odd integer that is large enough so that aliasing is

negligible. This leads to the following discretization of the spatial frequency variable

Q. The discrete spatial frequency is defined as $\mathbf{Q}_{mn} \equiv (\beta_m, \beta_n)$, where β_n is

$$\begin{aligned}\beta_n &= \frac{2\pi}{L}n, \quad 0 \leq n \leq \frac{M-1}{2} \\ &= \frac{2\pi}{L}[n-M], \quad \frac{M-1}{2} + 1 \leq n \leq M-1.\end{aligned}\tag{12}$$

Note that $\beta_{M-n} = -\beta_n$.

In this discussion the semiclassical model for photoelectric detection is used, i.e., the electromagnetic field incident on the detector is treated classically, but the interaction with the detector is quantized in units of photoevents.¹⁹ The intensities $I_{\mathbf{T}}(\boldsymbol{\rho}, z_d)$ and $I_{\mathbf{I}}(\boldsymbol{\rho}, z_d)$ are considered to be deterministic quantities. The actual number of photons counted in the mn^{th} pixel (centered at transverse coordinate $\boldsymbol{\rho}_{mn}$) is represented by the random variable N_{mn} . The expected number of photons is $\overline{N_{mn}} = W I_{\mathbf{T}}(\boldsymbol{\rho}_{mn}, z_d)$, where $W = (\Delta a \Delta t) \hbar c k_0$, and the overbar is the expectation operator, Δa , Δt , and \hbar are the effective pixel area (including the quantum efficiency), the measurement time, and Planck's constant divided by 2π , respectively.

With this scaling, $N_0 \equiv W E_{\alpha}^r \cdot E_{\alpha}^{r*}$, which is the expected number of reference beam counts from the first term in Eq. (7). It is convenient to define the random variable $c_{mn} \equiv N_{mn} - N_0$, and thus $\overline{c_{mn}} = W I_{mn}$, where I_{mn} is the holographic intensity from Eq. (8).

It is clear from Eq. (10) that unambiguous determination of $\tilde{\eta}$ without the advantage of prior information (other than finite support) requires at least two measurements with different experimental parameters. Let those two measurements be denoted (\mathbf{c}_1 and \mathbf{c}_2). The estimate of $\tilde{\eta}$ is represented by $\hat{\eta}$, which is to be determined

by the simultaneous solution of the Eq. (10) for the two data sets with $\tilde{I}(\mathbf{Q}_{mn})$ on the left hand side replaced by a measurement \tilde{c}_{mn}/W . Due to noise, $\hat{\eta} \neq \tilde{\eta}$. Regardless of the probability density function under consideration, $\tilde{I}_{mn} = \overline{\tilde{c}_{mn}}/W$, and Eq. (10) for the case of the two measurements takes the form:

$$\begin{aligned} \begin{pmatrix} (\tilde{c}_1)_{mn} \\ \overline{(\tilde{c}_1)_{mn}} \end{pmatrix} &= WH_1(\mathbf{Q}_{mn}) \begin{pmatrix} \hat{\eta}(\mathbf{Q}_{mn} + \mathbf{q}_{i1} - \mathbf{q}_{r1}) \\ \tilde{\eta}(\mathbf{Q}_{mn} + \mathbf{q}_{i1} - \mathbf{q}_{r1}) \end{pmatrix} \\ &\quad + WH_1^*(\mathbf{Q}_{M-m, M-n}) \begin{pmatrix} \hat{\eta}^*(\mathbf{Q}_{M-m, M-n} + \mathbf{q}_{i1} - \mathbf{q}_{r1}) \\ \tilde{\eta}^*(\mathbf{Q}_{M-m, M-n} + \mathbf{q}_{i1} - \mathbf{q}_{r1}) \end{pmatrix} \quad (13) \\ \begin{pmatrix} (\tilde{c}_2)_{p+m, q+n} \\ \overline{(\tilde{c}_2)_{p+m, q+n}} \end{pmatrix} &= WH_2(\mathbf{Q}_{p+m, q+n}) \begin{pmatrix} \hat{\eta}(\mathbf{Q}_{p+m, q+n} + \mathbf{q}_{i2} - \mathbf{q}_{r2}) \\ \tilde{\eta}(\mathbf{Q}_{p+m, q+n} + \mathbf{q}_{i2} - \mathbf{q}_{r2}) \end{pmatrix} \\ &\quad + WH_2^*(\mathbf{Q}_{p+M-m, q+M-n}) \begin{pmatrix} \hat{\eta}^*(\mathbf{Q}_{p+M-m, q+M-n} + \mathbf{q}_{i2} - \mathbf{q}_{r2}) \\ \tilde{\eta}^*(\mathbf{Q}_{p+M-m, q+M-n} + \mathbf{q}_{i2} - \mathbf{q}_{r2}) \end{pmatrix} \end{aligned}$$

where the integers $(p + M - m)$ and $(q + M - n)$ are to be calculated modulo M . The integers p and q are chosen so that the arguments of $\hat{\eta}$ and $\hat{\eta}^*$ in Eqs. (13) are equal to within the discretization of β given in Eq. (12). Thus, $(p, q) \approx (L/2\pi) [\mathbf{q}_{i1} - \mathbf{q}_{r1} - (\mathbf{q}_{i2} - \mathbf{q}_{r2})]$. Note that in the case of PSTM $\mathbf{q}_i = \mathbf{q}_r$ and $p = q = 0$. In order to simplify the notation, the equations (13) will be rewritten as follows, with the obvious symbol replacements.

$$\begin{aligned} \begin{pmatrix} c_1 \\ \overline{c_1} \end{pmatrix} &= WH_1^+ \begin{pmatrix} \hat{\eta}_+ \\ \tilde{\eta}_+ \end{pmatrix} + WH_1^{-*} \begin{pmatrix} \hat{\eta}_-^* \\ \tilde{\eta}_-^* \end{pmatrix} \\ \begin{pmatrix} c_2 \\ \overline{c_2} \end{pmatrix} &= WH_2^+ \begin{pmatrix} \hat{\eta}_+ \\ \tilde{\eta}_+ \end{pmatrix} + WH_2^{-*} \begin{pmatrix} \hat{\eta}_-^* \\ \tilde{\eta}_-^* \end{pmatrix} \end{aligned} \quad (14)$$

A. Gaussian PDF

In this section the probability density function (PDF) of the inverse discrete Fourier transform (IDFT) of the spatial distribution of photocounts measured on the detector is determined. Consider a Gaussian noise model in which each value of N_{mn} is statistically independent of the others and has a variance of σ^2 . The equations are made more transparent by defining the zero-mean random variable $b_{mn} \equiv N_{mn} - WI_{Tmn} = c_{mn} - WI_{mn}$. It is convenient to represent all of the values of b_{mn} as a single vector $\mathbf{b} = \{b_{mn}\}$, which will be called the *zero-mean photocount vector*.

Since the photocounts in each pixel are statistically independent of each other, the PDF of \mathbf{b} is:

$$P_{\mathbf{B}}(\mathbf{b}) = \frac{1}{(\sqrt{2\pi\sigma^2})^{M^2}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{n=0}^{M-1} \sum_{m=0}^{M-1} b_{mn}^2 \right\}. \quad (15)$$

The vector \mathbf{b} may be written in terms of its inverse DFT $\tilde{\mathbf{b}}$:

$$b_{mn} = \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} \tilde{b}_{kl} \exp \left\{ -i2\pi \left[\frac{km}{M} + \frac{ln}{M} \right] \right\} \quad (16)$$

Following the standard procedure for changing variables,¹⁹ Eq. (15) may be used to compute the PDF as a function of $\tilde{\mathbf{b}}$. Substituting Eq. (16) into Eq. (15) gives a sextuple sum. Making use of the identity:

$$\sum_{m=0}^{M-1} \exp \left\{ -i2\pi \frac{m}{M} (k+p) \right\} = M\delta(k - (M-p)), \quad (17)$$

where δ represents the Krönecker δ , Eq. (15) becomes:

$$\ln P_{\tilde{\mathbf{B}}}(\tilde{\mathbf{b}}) = \frac{-M^2}{2\sigma^2} \sum_{n=0}^{M-1} \sum_{m=0}^{M-1} \tilde{b}_{mn} \tilde{b}_{M-m, M-n} + \text{const}. \quad (18)$$

The PDF in Eq. (18) may be simplified. Since \mathbf{b} is real and the n^{th} and $(M-n)^{\text{th}}$ points are conjugate points in the M -point DFT, one it is seen that $b_{M-m, M-n} = b_{mn}^*$.

This relation allows one to reduce the number of terms in the sum in Eq. (18) by almost half when it is written in terms of a minimal parameter set as follows:

$$\ln P_{\tilde{\mathbf{B}}}(\tilde{\mathbf{b}}) = \frac{-M^2}{2\sigma^2} \left[\tilde{b}_{00}\tilde{b}_{00}^* + 2 \sum_{m=1}^{\frac{M-1}{2}} (\tilde{b}_{0m}\tilde{b}_{0m}^* + \tilde{b}_{m0}\tilde{b}_{m0}^*) + 2 \sum_{n=1}^{M-1} \sum_{m=1}^{\frac{M-1}{2}} \tilde{b}_{mn}\tilde{b}_{mn}^* \right] + \text{const.} \quad (19)$$

Thus, each Fourier component included in this minimal parameter set is statistically independent of the others. If the variance of N_{mn} were not constant with m and n , the various Fourier components would be coupled in the PDF.

When there is more than one measurement of the photocount vector, presumably each measurement is statistically independent and therefore the joint PDF for all measurements is simply the products of the individual ones. If \mathbf{b}_1 is the photocount vector of one measurement and \mathbf{b}_2 is that of another measurement, the joint PDF for the two measurements is given by:

$$P_{\mathbf{BB}}(\mathbf{b}_1, \mathbf{b}_2) = P_{\mathbf{B}}(\mathbf{b}_1)P_{\mathbf{B}}(\mathbf{b}_2) \quad \text{and} \quad (20)$$

$$P_{\tilde{\mathbf{B}}\tilde{\mathbf{B}}}(\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2) = P_{\tilde{\mathbf{B}}}(\tilde{\mathbf{b}}_1)P_{\tilde{\mathbf{B}}}(\tilde{\mathbf{b}}_2). \quad (21)$$

Eqs. (14) and (19) may be used to characterize quantitatively how much information is may be obtained as a function of spatial frequency \mathbf{Q} (in the object space), detection plane distance z_d , incident beam transverse wave vector \mathbf{q}_i , incident field \mathbf{e}_i , reference beam transverse wave vector \mathbf{q}_r and reference field vector \mathbf{e}_r .

Since $\tilde{b}_{mn} = \tilde{c}_{mn} - W\tilde{I}_{mn}$, Eqs. (14) and Eq. (21) yield

$$\begin{aligned} \ln P_{mn}(\hat{\eta}_-, \hat{\eta}_+^*, \hat{\eta}_-, \hat{\eta}_+) &= \frac{-W^2 M^2}{\sigma^2} \left\{ |H_1^+(\hat{\eta}_+ - \hat{\eta}_+) + H_1^{*-}(\hat{\eta}_-^* - \hat{\eta}_-^*)|^2 \right. \\ &\quad \left. + |H_2^+(\hat{\eta}_+ - \hat{\eta}_+) + H_2^{*-}(\hat{\eta}_-^* - \hat{\eta}_-^*)|^2 \right\} + \text{const.} \quad (22) \end{aligned}$$

In Eq. (22) $\hat{\eta}_-, \hat{\eta}_+^*, \hat{\eta}_-^*, \hat{\eta}_+$ are treated as independent variables. Eq. (22) is an example of the more general normal probability distribution $\ln P(\mathbf{x}) = \text{const} - (1/2)\mathbf{x}^H \mathbf{C}^{-1} \mathbf{x}$ where \mathbf{C} is the covariance matrix of the random vector variable \mathbf{x} and the superscript H is the Hermitian conjugate operator. The inverse covariance matrix in Eq. (22) is seen to be:

$$(\text{Cov})^{-1} = \frac{W^2 M^2}{\sigma^2} \begin{bmatrix} 0 & 0 & \varrho & \varepsilon \\ 0 & 0 & \varepsilon^* & \varsigma \\ \varrho & \varepsilon^* & 0 & 0 \\ \varepsilon & \varsigma & 0 & 0 \end{bmatrix}, \quad (23)$$

where $\varrho = H_1^- H_1^{-*} + H_2^- H_2^{-*}$, $\varepsilon = H_1^+ H_1^{-*} + H_2^+ H_2^{-*}$, and $\varsigma = H_1^+ H_1^{+*} + H_2^+ H_2^{+*}$. Eq. (23) has a symmetric block anti-diagonal structure. Its inverse is given by:

$$(\text{Cov}) = \frac{\sigma^2}{M^2 W^2 (\varrho \varsigma - \varepsilon \varepsilon^*)} \begin{bmatrix} 0 & 0 & \varsigma & -\varepsilon \\ 0 & 0 & -\varepsilon^* & \varrho \\ \varsigma & -\varepsilon^* & 0 & 0 \\ -\varepsilon & \varrho & 0 & 0 \end{bmatrix} \quad (24)$$

The variance of the estimated quantities, $\overline{|\hat{\eta}_+ - \tilde{\eta}_+|^2}$ and $\overline{|\hat{\eta}_- - \tilde{\eta}_-|^2}$, is given by the (2,4) and (1,3) elements, respectively, of the covariance matrix in Eq. (24). Since $\eta_+(\mathbf{Q}) = \eta_-(-\mathbf{Q})$, only one of these is non-redundant. Thus the variance of the estimate is given by,

$$\overline{|\hat{\eta}_+ - \tilde{\eta}_+|^2} = \frac{\varrho \sigma^2}{W^2 M^2 (\varrho \varsigma - \varepsilon \varepsilon^*)}. \quad (25)$$

In addition, the covariances of the real and imaginary parts of $\hat{\eta}_+$ and $\hat{\eta}_-$ may be determined from Eq. (24). Since the diagonal elements of the covariance matrix (24) are zero, the covariance of the real and imaginary parts are equal, and the real and

imaginary parts are uncorrelated. Thus the covariances of the real and imaginary parts of $\hat{\eta}_+$ and $\hat{\eta}_-$ are equal to 1/2 of the values on the right-hand side of Eq. (25). It is interesting to note that according to Eq. (24) $\hat{\eta}_+$ and $\hat{\eta}_-$ are correlated.

B. Poisson distribution case

Although determining the PDF of the estimators $\hat{\eta}_-, \hat{\eta}_+^*, \hat{\eta}_-, \hat{\eta}_+$ [as in equation (22)] in an analytically tractable way for the Poisson statistics case is challenging, determining the covariance matrix of the estimators as in Eq. (25) is relatively straightforward. Using the fact that $\overline{(N_{mn} - \overline{N_{mn}})^2} = \overline{N_{mn}}$ for a Poisson distributed random variable N_{mn} , one may show that:

$$\begin{bmatrix} \overline{c_1 c_1^*} - \overline{c_1} \overline{c_1^*} & \overline{c_1 c_2^*} - \overline{c_1} \overline{c_2^*} \\ \overline{c_2 c_1^*} - \overline{c_2} \overline{c_1^*} & \overline{c_2 c_2^*} - \overline{c_2} \overline{c_2^*} \end{bmatrix} = \begin{bmatrix} M^{-4} \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \overline{N_{1,mn}} & 0 \\ 0 & M^{-4} \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \overline{N_{2,mn}} \end{bmatrix}. \quad (26)$$

The equality associated with the (1,1) and (2,2) elements in Eq. (26) is essentially identical to the result given by Goodman¹⁹ in his discussion of stellar speckle interferometry (the M^{-4} factor being due to the difference between the IDFT and the DFT).

One may use Eq.s (14) to express the variance $\overline{|\hat{\eta}_+ - \tilde{\eta}_+|^2} = \overline{\hat{\eta}_+ \hat{\eta}_+^*} - \tilde{\eta}_+ \tilde{\eta}_+^*$ in terms of the quantities on the left hand side of Eq. (26). The result is:

$$\begin{aligned} \overline{\hat{\eta}_+ \hat{\eta}_+^*} - \tilde{\eta}_+ \tilde{\eta}_+^* &= \frac{1}{W^2(\rho_S - \varepsilon \varepsilon^*)} \\ &\times \left(\frac{|H_2^-|^2}{M^4} \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \overline{N_{1,mn}} + \frac{|H_1^-|^2}{M^4} \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \overline{N_{2,mn}} \right) \end{aligned} \quad (27)$$

In order to bring Eqs. (25) and (27) into agreement, one must have $N_0 = \sigma^2$, and $\sum_m \sum_n \overline{N_{1,mn}} = \sum_m \sum_n \overline{N_{2,mn}} = M^2 N_0$, which will always be true when $I_{Tmn} \gg I_{mn}$ (i.e., the object creates a weak hologram) and will sometimes be true otherwise. Then, Eq. (27) becomes:

$$\overline{\hat{\eta}_+ \hat{\eta}_+^*} - \tilde{\eta}_+ \tilde{\eta}_+^* = \frac{\rho N_0}{W^2 M^2 (\varrho \varsigma - \varepsilon \varepsilon^*)}. \quad (28)$$

C. Arbitrary noise case

It is instructive to consider a measure of the expected error independent of the particular noise model, so long as the noise is independently distributed (as was assumed in both cases above). One such measure is the squared l_2 norm of the system matrix inverse implied in Eq. (14).²⁰ That is, consider the matrix

$$A = \begin{bmatrix} H_1^+ & H_1^{-*} \\ H_2^+ & H_2^{-*} \end{bmatrix}, \quad (29)$$

and the vector

$$\hat{\boldsymbol{\eta}} = \begin{pmatrix} \hat{\eta}_+ - \tilde{\eta}_+ \\ \hat{\eta}_-^* - \tilde{\eta}_-^* \end{pmatrix}, \quad (30)$$

related to the data through Eq. (14), $\hat{\boldsymbol{\eta}} = A \tilde{\mathbf{b}}$. The error in the estimate is necessarily bounded by the error in the data and the norm of the matrix inverse. Explicitly

$$\|\Delta \hat{\boldsymbol{\eta}}\| \leq \|A^{-1}\| \|\Delta \tilde{\mathbf{b}}\|. \quad (31)$$

Consequently, the more singular A , i.e. the more poorly conditioned the system, the larger the magnification of the error in the estimate is expected to be. It is thus useful to consider $\|A^{-1}\|^2$ as an upper bound on the normalized error in the estimate.

4. Numerical results

A standard mathematics package (MATLAB) was used to evaluate numerically Eq. (25) for PSTM and TIRM. In the examples shown here the variance of the estimate was calculated as a function of spatial frequency at five heights $z_d = (0.03, 0.1, 0.3, 1.0, 10.0)\lambda$ (where $\lambda = 2\pi/k_0$). The measurements at greatest height being effectively of the TIRM type rather than PSTM. In PSTM the reference field is the evanescent incident field, so that $\mathbf{q}_r = \mathbf{q}_i$ and $\mathbf{e}^r = \mathbf{e}^i$. For the case that the field is measured at a detector height of 10λ , it is assumed that the reference field is still of the same polarization and transverse wave vector as the incident field. Two independent polarizations of the incident field are considered, TE ($\mathbf{e}^i \parallel \mathbf{k}(\mathbf{q}_i) \times \hat{\mathbf{z}}$) and TM ($\mathbf{e}^i \parallel \mathbf{k}(\mathbf{q}_i) \times \mathbf{k}(\mathbf{q}_i) \times \hat{\mathbf{z}}$) modes with the following parameters: $n = 2.0$, $\mathbf{q}_{i1} = (-2.0k_0, 0)$, and $\Delta z = 10^{-4}\lambda$. For one set of examples $\mathbf{q}_{i2} = (2.0k_0, 0)$, while for the second $\mathbf{q}_{i2} = (0, 2.0k_0)$. Thus the reference beams are anti-parallel for the first and perpendicular for the second. When $\mathbf{q}_{i2} = (2.0k_0, 0)$ the experiment achieves the maximum possible transverse resolution possible (in the x -direction).

Figures 1 and 2 show the variance of the estimate $\hat{\eta}(\mathbf{Q})$ as a function of spatial frequency. These “variance images” were normalized so that the smallest variance in any image is unity. The normalized variance images had a threshold set at 10^4 (i.e. values greater than 10^4 were set to that value) to reduce the dynamic range required of the display. Note that a value of 10^4 corresponds to an RMS error of 100 times that of the best determined value.

It may be seen from Figure 3 that the squared norm of the matrix inverse is a good

indicator of which regions of the data space will be practically accessible. However it should be noted that this upper bound does give quantitatively different results from those obtained for the variance with particular noise models, and so $\|A^{-1}\|^2$ cannot simply be taken as a substitute for the variance, nor can the noise model be ignored entirely.

It can be seen in Figures 1, 2 and 3 that at $z_d = \lambda$ the variance and the norm of the matrix inverse increase quickly above the far-field cutoff $|\mathbf{q}_i - \mathbf{Q}| < k_0$. The result indicates that the so-called far zone (the region where the evanescent fields are negligible) begins in the neighborhood of a single wavelength from the sample. If one hopes to take practical advantage of the high spatial frequency content of the near-field, measurements of the scattered field must be made within one wavelength of the sample. It should be noted that the shift of the Ewald sphere (or circle in this case) of reflection by the wave vector of the incident field remains evident even in the far zone. This shift may be used to advantage to circumvent the usual Abbe-Rayleigh limit when multiple incident fields of different wave vectors are used to sweep out a region of the Fourier space of the object outside the usual limiting circle of radius $2k_0$.

The case of an incident wave that is less evanescent, $\mathbf{q}_i = (-1.1k_0, 0)$, was also considered, though figures are not shown here. The results were qualitatively similar to the above case and led to similar conclusions. That is, at $z_d = \lambda$ the variance of the estimate increases quickly above the far-field cutoff [in this case $|\mathbf{Q}| < (1.1 + 1)k_0$]. Of course, in such a case the incident wave does not provide much more resolution than a homogeneous wave with $|\mathbf{Q}| = k_0$.

5. Conclusion

An analysis of achievable resolution for optical near-field microscopy with has been presented. An error in the estimate of the computed dielectric susceptibility was obtained for Gaussian and Poisson noise models and a model independent measure of expected error was similarly obtained. The formulae presented here place limits on the spatial resolution that a given experiment may be expected to achieve for thin (or at least invariant along the z -axis), weakly scattering samples.

The error in the estimate of the object structure was considered for measurements made at different heights above the sample. The region of the Fourier space of the sample practically accessible from these measurements is seen to decrease in size as the measurement plane is withdrawn to the far-zone. It was shown that the likely susceptibility estimate error increases dramatically for spatial frequencies above the far-field cutoff ($|\mathbf{q}_t - \mathbf{Q}| = k_0$), when the probe is even one wavelength away from the sample, indicating that the field does not contain much more information than is available in the far-field. The shift in the Ewald circle of reflection by the wave vector of the incident field was seen to be manifest even in the far zone indicating that superresolved imaging based on measurements taken in the far-zone alone is practically possible so long as the incident field is evanescent.

Objects exhibiting complicated structure, not separable in the transverse and longitudinal coordinates, are also of great interest. In those cases, one may employ similar methodology and make use of the inverse scattering formalism described in Ref. 13-16 in combination with a noise model to determine the likely errors in the

object reconstruction. This is the subject of a forthcoming paper.

Acknowledgments

We would like to thank J.C. Schotland for useful discussions. PSC wishes to acknowledge support from the US National Aeronautics and Space Administration under Grant NAG3-2764 and the National Science Foundation under CAREER grant 0239265.

Appendix

Expressions for the 1/2-space Green's tensor may be found in the literature.^{1,16} The tensor $\Gamma_{\alpha\beta}$ in Eq. (11) is obtained by integration of the Green's tensor over the longitudinal coordinate of the object (z') and is given by the expression:

$$\Gamma(\mathbf{q}) = \frac{1}{|\mathbf{q}|^2} \begin{pmatrix} q_x^2 h_{xx} + q_y^2 h_{yy} & q_x q_y (h_{xx} - h_{yy}) & |\mathbf{q}| q_x h_{xz} \\ q_x q_y (h_{xx} - h_{yy}) & q_x^2 h_{xx} + q_y^2 h_{yy} & |\mathbf{q}| q_y h_{xz} \\ |\mathbf{q}| q_x h_{zx} & |\mathbf{q}| q_y h_{zx} & |\mathbf{q}|^2 h_{zz} \end{pmatrix}, \quad (\text{A1})$$

where $\mathbf{q} = (q_x, q_y)$. The h functions are the result of multiplying the \tilde{g} functions in Ref. 16 by $\exp[i(k_z(\mathbf{q}_i) - k_z(\mathbf{q}))z']$ and integrating from 0 to Δz . Explicitly,

$$\begin{aligned} h_{xx} &= \frac{k_z^2(\mathbf{q})}{k_0^2} (\theta_1 + R_2 \theta_2) \\ h_{xz} &= \frac{-|\mathbf{q}| k_z(\mathbf{q})}{k_0^2} (\theta_1 - R_2 \theta_2) \\ h_{yy} &= (\theta_1 + R_1 \theta_2) \\ h_{yz} &= \frac{-|\mathbf{q}| k_z(\mathbf{q})}{k_0^2} (\theta_1 + R_2 \theta_2) \\ h_{zz} &= \frac{|\mathbf{q}|^2}{k_0^2} (\theta_1 - R_2 \theta_2), \end{aligned} \quad (\text{A2})$$

where $R_1(\mathbf{q})$ and $R_2(\mathbf{q})$ are the reflection coefficients given by

$$R_1(\mathbf{q}) = \frac{k_z(\mathbf{q}) - k'_z(\mathbf{q})}{k_z(\mathbf{q}) + k'_z(\mathbf{q})}, \quad (\text{A3})$$

and

$$R_2(\mathbf{q}) = \frac{k'_z(\mathbf{q}) - nk_z(\mathbf{q})}{k'_z(\mathbf{q}) + nk_z(\mathbf{q})}, \quad (\text{A4})$$

with $k'_z(\mathbf{q}) = \sqrt{n^2 k_0^2 - |\mathbf{q}|^2}$. The functions θ_1 and θ_2 are given by:

$$\begin{aligned} \theta_1(\mathbf{q}) &= i \frac{1 - \exp(i[k_z(\mathbf{q}_i) - k_z(\mathbf{q})]\Delta z)}{k_z(\mathbf{q}_i) - k_z(\mathbf{q})} \\ \theta_2(\mathbf{q}) &= i \frac{1 - \exp(i[k_z(\mathbf{q}_i) + k_z(\mathbf{q})]\Delta z)}{k_z(\mathbf{q}_i) + k_z(\mathbf{q})}. \end{aligned} \quad (\text{A5})$$

References

1. J-J Greffet and R Carminati, "Image formation in near-field optics," *Prog. Surf. Sci.* **56**, 133-237 (1997).
2. D. Courjon, K. Sarayedine and M. Spajer, "Scanning tunneling optical microscopy," *Opt. Comm.* **71**, 23-28 (1989).
3. M. Born and E. Wolf, *Principles of Optics*, 7th (expanded) edition (Cambridge, Cambridge Univ. Press, 1999).
4. E. Synge, "A suggested Method for extending Microscopic Resolution into the Ultra-Microscopic Region," *Phil. Mag.* **6**, 356-362 (1928).
5. E. Ash and G. Nicholls, "Super-resolution aperture scanning microscope," *Nature*, **237**, 510-512 (1972).
6. N. Garcia and M. Nieto-Vesperinas, "Direct solution to the inverse scattering problem for surfaces from near-field intensities without phase retrieval," *Opt. Lett.*, **20**, 949-951 (1995).

7. A. Lewis, M. Isaacson, A. Harootunian and A. Muray, "Development of a 500 Å spatial resolution light microscope. I. Light is efficiently transmitted through $\lambda/16$ diameter apertures," *Ultramicroscopy*, **13**, 227-231 (1984).
8. E. Betzig and J.K. Trautman, "Near-field optics: microscopy, spectroscopy, and surface modification beyond the diffraction limit," *Science*, **257**, 189-195 (1992).
9. C.W. McCutchen, "Optical systems for observing surface topography by frustrated total internal reflection and interference," *Rev. Sci. Instr.*, **35**, 1340-1345 (1964).
10. P.A. Temple, "Total internal reflection microscopy: a surface inspection technique," *Appl. Opt.*, **20**, 2656-2664, (1981).
11. D. Courjon, J-M Vigoureux, M. Spajer, K. Sarayedine and S. Leblanc, "External and internal reflection near field microscopy: experiments and results," *Appl. Opt.* **29**, 3734-3740 (1990).
12. P.T.C. So, H.-S. Kwon and C. Dong, "Resolution enhancement in standing-wave total internal reflection microscopy: a point-spread-function engineering approach," *Journ. Opt. Soc. Am. A*, **18**, 2833-2845 (2001).
13. P.S. Carney and J.C. Schotland, "Inverse scattering for near-field microscopy," *Appl. Phys. Lett.* **77**, 2798-2800 (2000).
14. P.S. Carney and J.C. Schotland, "Three-dimensional total internal reflection microscopy," *Appl. Phys. Lett.* **26**, 1072-1074 (2001).
15. P.S. Carney and J.C. Schotland, "Determination of three-dimensional structure in photon scanning tunneling microscopy," *J. Opt. A: Pure Appl. Opt.* **4**,

S140-S144 (2002).

16. P.S. Carney and J.C. Schotland, "Theory of total-internal-reflection tomography," J. Opt. Soc. Am. A **20**, 542-547 (2003).
17. D.G. Fischer, "The information content of weakly scattered fields: implications for near-field imaging of three-dimensional structures," J. Mod. Opt. **47**, 1359-1374 (2000).
18. J.W. Goodman, *Introduction to Fourier Optics*, 2nd edition (McGraw-Hill, New York, 1988) pp. 22-27.
19. J.W. Goodman, *Statistical Optics* (Wiley, New York, 1985).
20. T.K. Moon and W.C. Sterling, *Mathematical Methods and Algorithms for Signal Processing* (Prentice-Hall, Upper Saddle River, NJ, 2000), pp. 140-141, 235-237.

Fig. 1. The base-10 logarithm of normalized PSTM susceptibility Fourier space variance for detection heights of $(0.03, 0.1, 0.3, 1.0, 10.0)\lambda$ as labeled above. The coordinates are spatial frequency in units of $k_0 = 2\pi/\lambda$. The data are taken to consist of two scans with incident wave vectors of $\pm 2k_0\hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is a unit vector. Results of illumination by TM modes are shown on the left and TE on the right. The plots are normalized so that the minimum variance of any point is 1.0. The linear color scale runs from 0 to 4 and is shown at the bottom of the figure. Note that the range on each figure is different. Values of the normalized variance greater than 10^4 were set equal to 10^4 for clarity in display. It can be seen that at $z_d = 1.0\lambda$ the effective information content is already similar to that of the far-field limit, in which only the homogeneous modes are detected.

Fig. 2. Same as Figure 1, except that the data are assumed to consist of two scans with incident wave vectors of $-2k_0\hat{\mathbf{x}}$ and $2k_0\hat{\mathbf{y}}$, where $\hat{\mathbf{y}}$ is a unit vector.

Fig. 3. The logarithm of the squared l_2 norm of the system matrix inverse for the cases of measurements made in the two closer planes. The four top plots show results for the case of counter propagating incident evanescent waves, while the bottom four plots show results for the case of the same orthogonal incident wave vectors discussed above. Coordinates are again spatial frequency in units of k_0 . The linear color scale running from 0 to 4 is shown at the bottom of the figure.

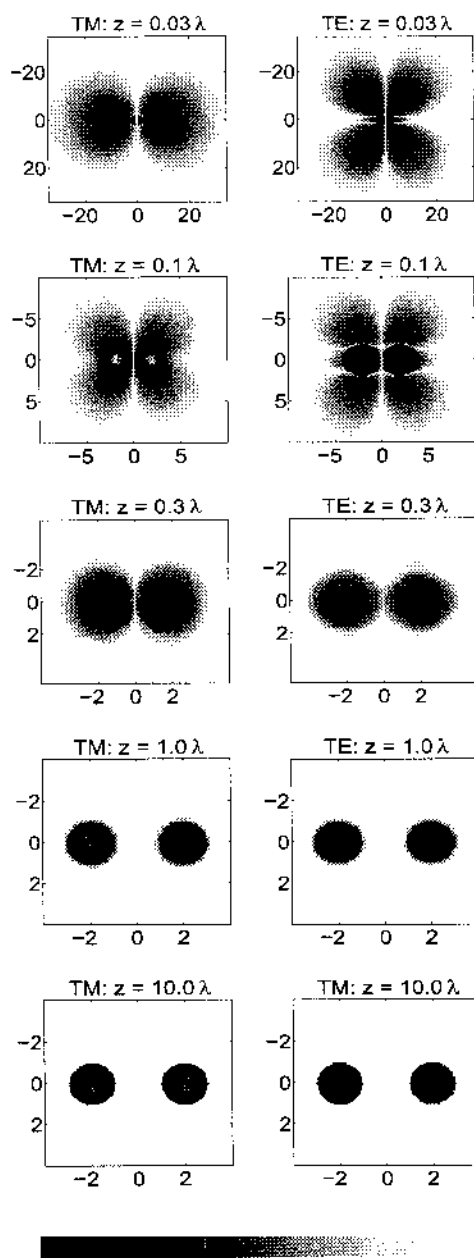
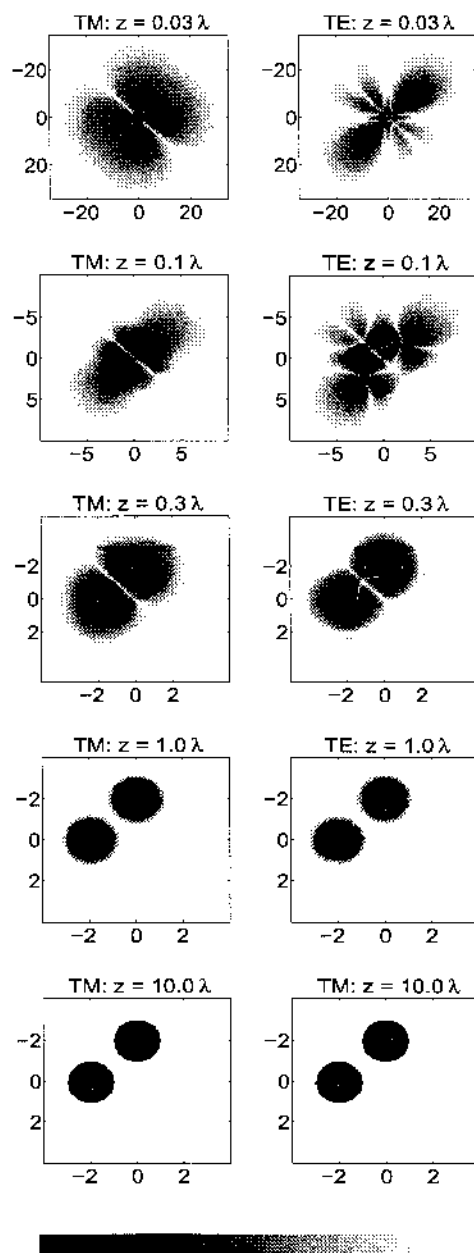


Fig. 1



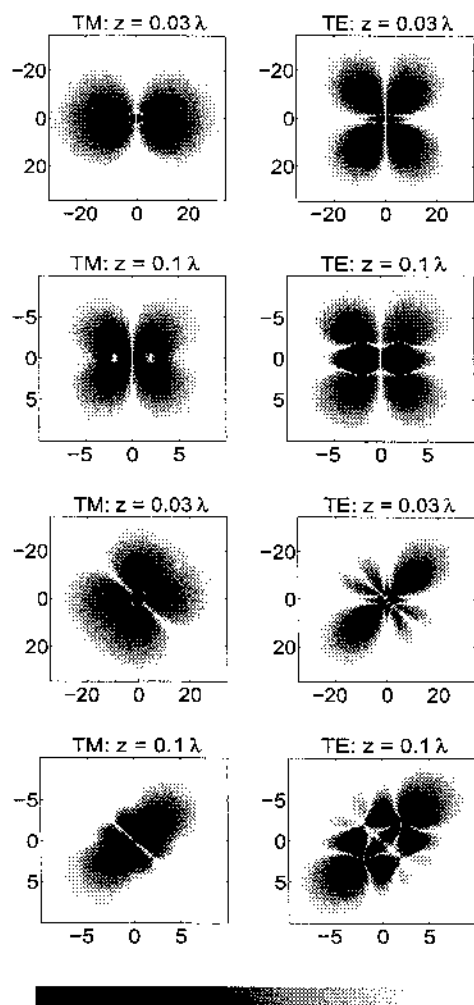


Fig. 3